13

Behavioral Game Theory

COLIN F. CAMERER

People deviate from the predictions of game theory in two systematic ways. They are not purely self-interested (they care about fairness and try to cooperate with others), and they do not always consider what other players will do before making choices. However, with experience, these deviations sometimes disappear. People learn when they can afford to be unfair and what others will do; their behavior often converges to a game-theoretic equilibrium. A behavioral game theory that explains the initial deviations (and their disappearance) could be useful, especially if the learning process is modeled carefully and better data are gathered.

BEHAVIORAL DECISION THEORY is a catalog of ways in which judgments and choices deviate from normative decision theory and of psychological explanation of these deviations. Despite the formal kinship between decisions and games, there is no behavioral *game* theory. In this chapter, I describe some data that suggest a basis for behavioral game theory.

My approach expands the simple way in which special features of games (as compared to decisions or competitive markets) are treated in normative game theory. Games have two special features: players might care about the payoffs others get, and players must make judgments about the choices others make (and about their own future choices, in dynamic games). In game theory, it is generally assumed that people are self-interested—they do not care about the payoffs of others—and use introspection to make accurate judgments about the choices of others (who are making simultaneous judgments by introspection, ad infinitum).

These assumptions are useful for deriving sharp equilibrium predictions. Without them, game theory is still quite useful as a system for classifying social situations (Aumann 1985). The important question is whether the as-

Thanks to Robyn Dawes, Howard Kunreuther, Keith Weigelt, and especially George Loewenstein and Robin Hogarth, for help. This research was supported by National Science Foundation grant SES87-08566.

sumptions are violated systematically enough for an alternative theory to be useful.

A wide array of evidence suggests that they are. Preferences are more complicated than simple self-interest, but they are highly context dependent. Judgments about choices of others are less complicated than introspective equilibrium calculations, but they converge to those calculations as people learn from feedback over time.

There are other useful ways to do behavioral game theory. One way is to ask how the lessons of behavioral decision theory apply in games. For instance, one can ask how outcomes of bargaining situations depend on the way negotiators frame outcomes. This approach has been taken successfully by others in bargaining (e.g., Bazerman and Carroll 1987; Mumpower 1988); I will not retrace their steps. Behavioral decision theory also points out psychological features, like the importance of context and conditions for learning, that are useful in understanding the empirical convergence of behavior to game-theoretic predictions. Another useful direction is to reexamine the normative status of game theory. In this reexaminations, behavioral considerations arise naturally from wondering how people think about games rather than from empirical evidence. Binmore (1987), Rubinstein (1988), and Fudenberg and Kreps (1988) are provocative this way.

NOTATION AND BASIC DEFINITIONS

A *game* consists of *players* ($i = 1, \ldots, n$); *strategies* that players choose (s_i for player i); *outcomes* that result from strategy choices, a function of the vector ($s_1, \ldots, s_n$); *preferences* that players have for outcomes, including lotteries over possible outcomes ($u_i\,[s_1, \ldots, s_n]$); and rules about the order of moves, the information players have at each point, and so on. I will discuss only games played *noncooperatively*, in which players cannot make binding agreements about what to choose. Whereas noncooperative game theory is concerned with the strategies that players choose, *cooperative* game theory is mostly concerned with the division of gains from the strategies that are chosen by binding agreement. A noncooperative game can be shown in a tree ("extensive form") or in a matrix ("strategic" or "normal form").

The obvious question in a noncooperative game is what strategies players will choose. Nash (1951) suggested that players might choose strategies that are best responses to each other. Such strategies form a *Nash equilibrium*. Formally, ($s^*, \ldots, s_n^*$) is a Nash equilibrium if and only if

$$u_i(s_1^*, \ldots, s_i^*, \ldots, s_n^*) \geqslant (s_1^*, \ldots, s_i, \ldots, s_n^*) \qquad (1)$$

for all s_i, i. Nash equilibrium is a simple solution concept with attractive properties: an equilibrium always exists for games with finitely many strategies and players, and it is easy to calculate and explain. Many other equilibrium concepts extend Nash's idea (e.g., Aumann 1987) or refine it (see below).

The plan of the chapter is as follows. First, experimental evidence about preferences—tastes for fairness and cooperation—is reviewed. Next, data about judgments of future choices and the choices of others are discussed. (These discussions are extremely selective, intended to illustrate arguments rather than provide comprehensive review. For more thorough reviews, see Roth [1987], Kahan and Rapoport [1984], and McKelvey and Ordeshook [1987] on cooperative games, Colman [1983] on experiments, and Aumann [1985] for wisdom.) Then, some parallels to behavioral decision theory are discussed. A final section offers conclusions and ideas for further research.

PREFERENCES

Fairness

People prefer payoffs that are fair. This is not inconsistent with game theory because players are assumed to have utility for an outcome, which produces a vector of payoffs (one for each player). A player's utility can certainly depend on the payoffs that others get. Assuming that there is no dependence—pure self-interest—is simply a convenient benchmark, like the assumption of risk aversion in risky choice. If we reject the self-interest assumption, we do not reject game theory.

In fact, we can reject the self-interest assumption. Furthermore, we cannot account for fairness preferences by simply assuming that people care about the payoffs of others because their caring depends idiosyncratically on context. Let us see some examples.

The Coase Theorem. The Coase theorem is the conjecture that socially efficient outcomes will result, regardless of who has the right to make decisions about imposition of economic "externalities," as long as people can bargain cheaply.[1]

Many experiments have tested this conjecture in very simple settings (see Hoffman and Spitzer 1982, 1985, 1986). In a typical experiment, a "controller" subject A chooses one of several possible divisions of money between A and B, for example, (4, 10), (6, 6), or (11, 0). (These divisions represent costs and benefits to two parties because of different economic externalities.) The controller and B bargain with each other in an unstructured way, but the controller has the right to choose any division she wishes if they disagree. Since the experimenter has made it easy for the players to bargain, the Coase

1. An externality is any effect one party has on another, good or bad, that lies outside their economic relationship. Examples include watching an attractive man or woman, hearing a baby scream on an airplane, or being trapped on a boat listening to a band you dislike. Coase (1960) argued that, if bargaining between parties is easy, the socially efficient outcome would result regardless of who has the "property right" to impose or prevent the externality.

theorem predicts that controllers will choose the division with the largest total payoff (the socially efficient outcome), then demand a payment from B. In the example, the self-interested controller should choose (4, 10), then demand a side payment of at least 7 to bring her total to 11 (because she could pick [11, 0] if she wanted to).

In the experiments, subjects almost always chose the efficient outcome, but controllers got adequate side payments in only a third of the cases; they often split the efficient outcome evenly (7, 7). Similar results are found with larger groups (Hoffman and Spitzer 1986), in a two-firm market setting (Prudencio 1982), and when the subjects bargain over payments for drinking a distasteful substance (Coursey, Hoffman, and Spitzer 1987). Even splits are common in many other bargaining situations too (Roth 1987).

Hoffman and Spitzer (1985) explored two ways to make subjects tolerate uneven splits: in a "moral authority" treatment, subjects were told that they "earned the right" to be controller; in a "game trigger" treatment, subjects became controller if they won a simple game of skill (Nim). Both changes in context led controller subjects toward more uneven allocations. (The moral authority treatment was stronger, which goes curiously unmentioned in many discussions of the results.) Harrison and McKee (1985) found that having several practice periods with no controller, which generally led to disagreement and inefficient outcomes, led to uneven splits in later periods. These data suggest that an unequal division can be acceptable to people if they think that the right to the larger share has been earned.

Ultimatum Games. More evidence of preferences for fair allocations comes from "ultimatum games." In an ultimatum game, the divider divides a $10 "pie" by keeping D and giving $\$10 - D$ to an accepter (written $[D, 10 - D]$). She accepts the division, and they get paid; or she rejects it, and they get nothing. (Ultimatum games are not common in life, but they are a lens through which attitudes toward others' payoffs can be seen. They also underlie more elaborate theories of bargaining, discussed below, that do apply to common situations.)

There are many Nash equilibria of the ultimatum game, but, in the most reasonable ("subgame perfect")[2] equilibria, the divider should leave a penny (or nothing) to the accepter. A purely self-interested accepter will pick up the penny, unenthusiastically.

People do not actually play ultimatum games so ruthlessly; they leave

2. Any division $(D, 10 - D)$ is a Nash equilibrium if the divider believes that the accepter would reject all other divisions. But such a belief seems paranoid and inconsistent: it assumes actions in parts of the game (subgames) that people would not actually take if those parts were reached. Such an equilibrium is not subgame perfect. The division (5, 5) is a Nash equilibrium, but it is not subgame perfect because it assumes that the accepter will reject (9, 1) (or any other division less favorable than [5, 5]) if the (9, 1) subgame is reached. If the (9, 1) subgame is reached, a self-interested accepter will accept rather than reject.

around 40 percent of the pie for the accepter. Splits leaving less than 20 percent to the accepter are often rejected, contrary to pure self-interest.[3] Auctioning off the right to divide does not change the results much (Güth and Tietz 1986).

Buying from a monopoly is like playing an ultimatum game (see also Thaler 1988, 202–3). For example, in a "posted offer" experiment, a single seller posts a price at which she will sell a good. Buyers should buy if their reservation price is above the offered price and reject the offer otherwise. (Since the price is posted and fixed, there is no room for haggling.) In experiments, buyers *sometimes* refused to buy, even when the good was worth more to them than it cost (Smith 1981; Coursey, Isaac, and Smith 1984). These buyers were either trying to force prices down or were punishing the monopolist for posting unfair prices. Their efforts resemble boycotts, like the one endorsed by New York City mayor Ed Koch to punish movie theaters for raising ticket prices to $7 in the 1980s. Boycotts inspired by fairness are probably *not* effective in the long run—Koch's was not—but they may prevent prices from adjusting rapidly to changes.

There is *corroborating* evidence of tastes for fairness or altruism from many sources. Selten (1987) reports evidence from cooperative game experiments. Loewenstein, Thompson, and Bazerman (1989) found that people disliked differences between their payoffs and others' payoffs (especially differences favoring others) and that their dislike was marginally decreasing. Many earlier studies found similar results.

While people are not purely self-interested, we cannot just assume that they have a utility for payoffs of others; whether they do depends on context (as the Coase theorem data show). Our attention must shift to precisely how context matters.

Surveys of hypothetical situations are one useful way to study the implicit rules that people have for relating fairness to context. The rules uncovered this way do not correspond closely to formal rules in game theory (Yaari and Bar-Hillel 1984) or economics (Kahneman, Knetsch, and Thaler 1986a, 1986b), except perhaps for M.B.A. students (Kunreuther 1986). For example, biological need is considered a fairer basis for a disproportionate claim than simple desire (Bar-Hillel and Yaari 1987). Rationing scarce objects by raising prices is considered less fair than making people wait in lines or win lotteries

3 Guth, Schmittberger, and Schwarze (1982) found that dividers left an average of 35 percent of the pie (with experience, 31 percent). In Kahneman, Knetsch, and Thaler (1986a), dividers left 45 percent of the pie; *most* of the divisions were equal splits. In classroom replications, my students left 39 percent of a $10 pie and 38 percent of a $100 pie. (One pie of each size was actually divided.)

Guth, Schmittberger, and Schwarze's inexperienced subjects rejected two of 21 divisions, which left an average of 10 percent to the accepter. Experienced subjects rejected six of 21, leaving an average of 22 percent. Kahneman, Knetsch, and Thaler's subjects rejected unless they got 23 percent. My students rejected unless they got 21 percent of $10 or 15 percent of $100.

for them. People do not object to wage freezes during times of inflation—which reduce real, inflation-adjusted wages—but they think that absolute wage decreases are unfair. These studies suggest that the context of economic transactions flavors their fairness in subtle, idiosyncratic ways.

Cooperation

The instinct to cooperate is another kind of preference that departs from strict self-interest. Cooperation is indicative of concern for others because it may result from an aversion to the unfairness that results from uncooperative choices. Some examples will illustrate the point.

Social Dilemmas and Public Goods. In a social dilemma, a person's contribution benefits the group more than it benefits her. (In a typical experiment, I can keep $5 or give it back to the experimenter, who distributes $1 each to 10 people.) She prefers to withhold the contribution ("free ride"), but, if everyone withholds, each individual is worse off. Many social situations, for example, bystander helping, resemble social dilemmas (Stroebe and Frey 1982). In economics, social dilemmas arise in the funding of "public goods," goods that can be supplied to additional people at low marginal cost and that people cannot be excluded from consuming (such as national defense or public art).

There have been many experimental studies of social dilemmas (mainly by sociologists and psychologists) and public goods (mainly by economists and political scientists). I will describe them very briefly. For more complete reviews, see Dawes (1980), Dawes and Orbell (1981), Messick and Brewer (1983), and Dawes and Thaler (1988).

The most basic finding is that subjects contribute more to public goods than they should according to pure self-interest. Initial contributions average around half the optimal level, then dwindle to 10–20 percent (Marwell and Ames 1979, 1980, 1981; Brubaker 1982; Isaac, McCue, and Plott 1985; Kim and Walker 1984; Harrison and Hirshleifer, in press). People who value the public good more highly or have more resources contribute more (e.g., Rapoport 1988).

Subjects who contribute initially may do so because they have not learned to free ride or because they are building reputations for cooperativeness, which induces cooperation in others (see the discussion of reputation games below). Andreoni (1988) noticed reputation building among players paired with the same group in a series of 10-period games. In the first 10-period game, contributions began at a high level in the first period and gradually dwindled. In the first period of a second 10-period game, contributions jumped up from the low level to which they had dwindled to a high level, as if some subjects were trying to rekindle cooperation.

Many experiments have studied step-level public goods, which yield increasing benefits to participants in discrete increments (e.g., a bridge or a fleet of ships). An elegant experimental paradigm for such goods was proposed by van de Kragt, Orbell, and Dawes (1983). Each of N players is given E dollars, which they can keep or contribute. If M players contribute (the "minimal contributing set"), the public good is supplied, and all players get R (with $R > E$). In step-level public goods, free riding is not a dominant strategy. If a player expects $M - 1$ others to contribute, then it pays for her to contribute, earning R instead of E.

Roughly half the players contribute to step-level public goods (e.g., van de Kragt et al. 1986; Rapoport 1988) if there is no discussion. (Discussion almost always enables groups to ensure that M players contribute; see van de Kragt, Orbell, and Dawes 1983). In the "no greed" game (Simmons, Dawes, and Orbell 1983), contributions are required from everyone if M people contribute. Contributions are much more frequent (about 90 percent) in this game; it seems that people do not contribute in regular games because they can both keep their endowment and share the public good.

The evidence against strong free riding is so overwhelming that thoughtful researchers have begun to study conditions under which more and less cooperation occurs. In social dilemmas, discussion about contributions works wonders, roughly doubling contributions from one- to two-thirds (Dawes, McTavish, and Shaklee 1977). A "sermon" by the experimenter increases contributions too (cf. the success of telethons). Group discussion about something other than the game and time to think about the dilemma do not help.

Orbell, van de Kragt, and Dawes (in press) found that subjects in a discussion group would contribute twice as often if they thought during the discussion that their contributions benefited members of their own group rather than members of a separate group (created minutes earlier by randomly dividing one large group into two). It seems that discussion creates group identity and loyalty quickly and persistently (Dawes, van de Kragt, and Orbell 1988); moreover, much weaker conditions do too (such as giving subjects a common random payoff; Kramer and Brewer 1986).

Cooperation in Bargaining. In cooperative games with incomplete information, players are assumed to have private information about their own values. The important question is how they will divide the gains from cooperating. Much of the theoretical research in this area centers around a formal equivalence—the "revelation principle"—between freewheeling bargaining among players and structured bargaining guided by an outside arbitrator (e.g., Myerson 1986). The theories usually predict that inefficiencies are necessary to keep players from lying about their information to the (mythical) arbitrator.

For instance, in Forsythe, Kennan, and Sopher's (1987) bargaining games with two players, one "informed" player knew the size of the pie to be split (e.g., either \$6 or \$1, equally likely). The two players had 10 minutes to decide how to split the pie; if they could not agree, they got nothing. To keep the informed player honest, there must be some penalty for claiming that the pie is small when it is actually large ("we do not have enough profit to pay for a wage increase"). The theoretical penalty is a chance of disagreement when the pie is small (akin to labor strikes); such disagreements are very costly when the pie is actually large. For example, suppose that the pie is $6 but that the informed player says that the pie is $1, offers a $.50–$.50 split, and provokes a disagreement. Then the opportunity to split $6 has been lost. If such disagreements are sufficiently likely, the informed player should not lie.

Another maintained assumption in the theory is that informed players should make the same offers whether the pie is $6 or $1. They did not. About a third of the time, when the informed player knew the pie was $6, she offered more than $1 to the other player (e.g., a $3–$3 split), immediately revealing that the pie was not $1. (Experience did not diminish the number of revealing offers.) Revealing offers were costly because informed players who made them ended up earning $1 less than informed players who were more inscrutable. However, by revealing the size of the pie, informed players provoked fewer disagreements than the theory predicted, which increased the total earnings of all subjects.

Radner and Schotter (1987) experimented with games in which a buyer and seller know their value and cost, respectively, and both players know the distributions of possible values and costs. Each player writes down a bid or ask price. If the bid and ask overlap, a trade takes place at a price midway between them; otherwise, there is no trade. (If the buyer bids $4 and the seller asks $2, a trade takes place at the price of $3.)

It is optimal for players to bid somewhat less than their values (e.g., bid $2 if the valuation is $3) and ask more than their cost. Behaving this way maximizes the expected gain from trading, but it causes inefficiencies because some trades that should take place do not. Subjects in experiments bid much closer to their true values than they should. This "irrationality" caused much less inefficiency than the theory predicted, as if subjects were cooperating to maximize their collective gains from trade (taking the most money from the experimenter). Face-to-face bargaining was the most efficient of all.

The pie-splitting and bid-ask data suggest that subjects are unwilling to conceal their private information completely (or are unable to do so because the "availability" of information in memory makes them think that others know it too; see Camerer, Loewenstein, and Weber 1989). Their behavior is cooperative because it requires personal sacrifice—revealing information hurts informed players—that benefits players collectively, just as contributing in the social dilemma does.

JUDGMENTS ABOUT FUTURE CHOICES IN DYNAMIC GAMES

In decisions, people make judgments about random events; in games, people must make judgments about the choices of others. In dynamic games, people must make judgments in early stages about what they will do later.

These judgments are initially myopic; people do not anticipate what will happen in future plays and use those anticipations to make good choices in early plays. However, in most experiments, people learn not to be myopic.

Sequential Bargaining

Rubinstein (1982) pioneered the study of alternating-offer sequential bargaining games in which player 1 makes an offer, player 2 accepts it or rejects it and makes a counteroffer (which player 1 accepts or rejects), and so on.

One can study the effect of bargaining costs by making the size of the pie shrink each period by a fixed percentage (representing impatience or a discount rate) or by a fixed amount (representing bargaining costs). For a review of theory in this area, see Sutton (1986).

The tendency toward fairness in ultimatum games led many experimenters to study these more complicated sequential bargaining games, which end in ultimatums. For example, Neelin, Sonnenschein, and Spiegel (1988) used 2-, 3-, and 5-period games (cf. Binmore, Shaked, and Sutton 1985, 1988). In the 3-period game, the pie sizes were \$5, \$2.50, and \$1.25. The perfect equilibrium prediction is that the player moving first should keep \$3.75 and offer \$1.25 to the second player (who should accept the offer).

(In the third period, an ultimatum game, the first player could demand \$1.25 in theory. Anticipating this, the second player could demand only \$1.25 of the \$2.50 pie in the second period; if she took more, the first player would refuse and move on to the third period to get \$1.25. Anticipating both stages, the first player can demand \$3.75 initially, leaving the second player \$1.25, the amount she could get by refusing and moving to the second stage.)

Subjects played very close to the perfect equilibrium prediction in the 2-period game, but they acted as if the 3- and 5-period games would only last 2 periods. For instance, in the 3-period game, subjects typically offered \$2.50 (50 percent of the pie) initially, as if the second period were an ultimatum game and the second player could certainly earn \$2.50 (though it was not). Perfect equilibrium also predicts badly in the comprehensive study by Ochs and Roth (1989) and in 2-period games in which the pie shrinks by 10 or 90 percent (Güth and Tietz, 1988).

In these experiments, subjects did not usually play the entire multiperiod game because first-period offers were accepted 80–90 percent of the time. In similar experiments on multiperiod assets, subjects underestimate the importance of future periods for current-period asset prices until they have actually lived through the future periods (e.g., Forsythe, Palfrey, Plott 1982). Knowing this, Harrison and McCabe (1988) replicated the 3-period game results of Neelin, Sonnenschein, and Spiegel with a clever twist.

Their subjects played an entire 3-period game, then played a separate 2-period game in which the pies were \$2.50 and \$1.25. This 2-period game is a subgame of the 3-period game with pies of \$5, \$2.50, and \$1.25: if the initial offer had been rejected in the 3-period game, the players would have found themselves playing two more periods equivalent to the separate 2-period game that they did play. After the separate 2-period game, they played another 3-period game, followed by another 2-period game, and so on.

Subjects initially split the \$5 equally in the 3-period game, just like the subjects in Neelin, Sonnenschein, and Spiegel's study. Then they played a 2-period game and saw that the first player ended up with \$1.25. Players moving first gradually realized that player 2 would get only \$1.25 in the 3-period game if she rejected the initial offer, so they gradually raised their initial demands to \$3.75, the perfect equilibrium prediction.

Perfect equilibrium in the 3-period game resulted from seeing the results of future periods (subgames) played out—from experiential backward induction—rather than from hypothetical backward induction.

In this particular game, experiential backward induction also taught them that they did not have to be fair (player 1 gets \$3.75 of the \$5 pie), though an equal split was a natural division to start with. A reasonable conjecture is that players in unfamiliar bargaining environments begin by behaving fairly. If the environment favors one person over another, they gradually learn of the advantage and exploit it (see Binmore, Shaked, and Sutton 1985).

Strategic and Sincere Voting

Experiential backward induction also teaches people to vote strategically in the presence of a voting agenda. The experiment of Eckel and Holt (1989), shown in figure 13.1, is a good example. Nine subjects vote on three alternatives, A, B, and C. Three subjects, called A voters, prefer A to C to B (written $A > C > B$) because the experimenter pays them \$3 if A is elected, \$2 if B is elected, and \$1 if C is elected. Three B voters have preferences $B > A > C$, and three C voters have $C > B > A$.

Subjects were given a fixed agenda. On the first vote, they chose either {A, B} or {B, C} (i.e., they would decide which one of A and C would later run against B). If {A, B} was voted in, on the second vote they chose A or B; if {B, C} was voted in, they chose B or C.

Players are said to vote "sincerely" if they vote for the set that contains their most preferred alternative. Under sincere voting, A voters would vote for {A, B} over {B, C}. B voters would too, so {A, B} would win. In the {A, B} runoff, A would get only three votes (from the A voters) and lose to B. It is myopic for A voters to choose {A, B} in the first stage because they will end up with B, which they like least of all. It is smarter for A voters to vote "strategically" by choosing {B, C} in the first stage (voting against their true preference), thereby setting up a runoff between B and C that C would win. (A voters would rather end up with C than B. Note that it does not pay for B or C voters to vote strategically.)

The entire 2-period game was repeated about 10 times in the experiment. No A voters voted strategically (for {B, C}) at first, but half their votes were strategic by the fifth repetition. Whether the preference orders of all subjects

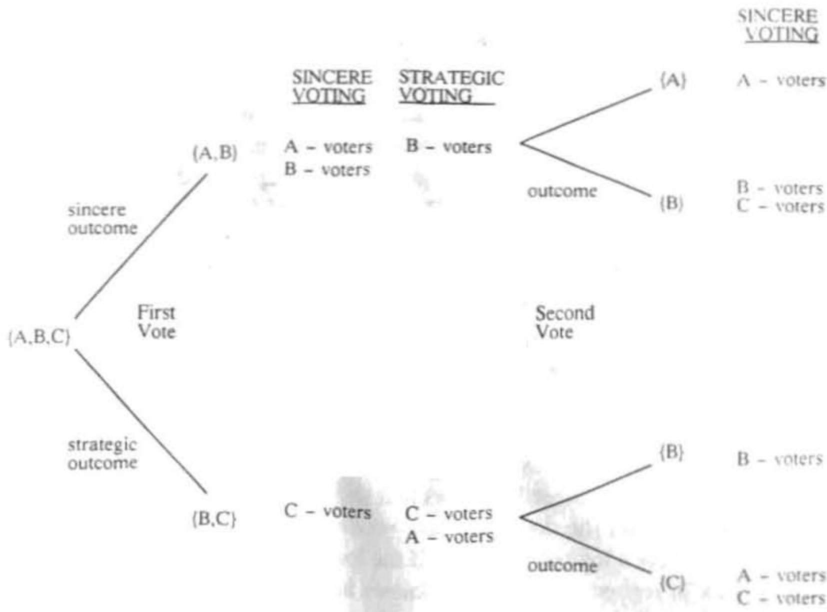

Figure 13.1. A voting game (Eckel and Holt 1989). Preferences: A voters, A > C > B; B voters, B > A > C; and C voters, C > B > A.

were made public before the voting did not matter much: as in the sequential bargaining experiments, subjects seemed to learn by living through future periods (experiential backward induction) rather than by thinking about them (hypothetical backward induction).

JUDGMENTS OF CHOICES BY OTHERS

Players in games must make judgments about choices of other players. (This is what distinguishes games from decisions most sharply.) The decision-theoretic approach treats choices by others as random events, like weather or disease (Kadane and Larkey 1982). The game-theoretic approach treats choices by others as special because their choices can be predicted (nature cannot) if we know what they earn, think they are rational, and believe they have the same presumption of us.

The crucial question is how much faith one has in human rationality. Harsanyi (1982, 121), a game theorist, objects to the decision-theoretic approach because it "would amount to *throwing away essential information*, viz., the assumption . . . that the players will act rationally and will also *expect* each other to act rationally." To call the mutual rationality assumption "information" seems curious (unless a mathematician is talking); an assumption is a hypothesis, subject to test. Game theory just provides a benchmark from which systematic deviations of the mutual rationality assumption are defined.

Reputation Games

Harsanyi (1967–68) suggested that games in which players have private information could be modeled by assuming that nature generates a player's "type," which affects payoffs and which is known only to that player (though others know the probabilities of various types). When played repeatedly, such games provide a natural model of the formation of reputations. Games like this are fashionable in economics and political science as models of labor and product markets, strikes, campaigns, and so on (Wilson 1985).

Camerer and Weigelt (1988) experimented with a reputation game consisting of eight plays of the 1-period ("stage") game shown in figure 13.2. The game is a simple model of building trust. First, nature determines the entrepreneur's (E) type, either honest ($p = .1$) or dishonest ($p = .9$). The banker (B) does not observe E's type. Her ignorance is shown in figure 13.2 by lumping the honest and dishonest nodes of the tree together in an information set enclosed by a dotted ellipse. B knows that she is at one of the two nodes in the information set, but she does not know which one.

B either offers a loan or does not. If the loan is offered, E chooses whether to pay back or renege. Notice that E knows her own type; she knows whether she is on the left (honest) part of the tree or the right (dishonest). If she is dishonest, she prefers reneging (earning 150) to paying back (earning 60). If she is honest, she prefers paying back, earning 60 instead of 0. A single E plays the game eight times against different Bs. Each B can observe what happened before. E knows her own type before the 8-period game, but B knows only the probability of each type.

The sequential equilibrium[4] is complicated. An honest-type E should always pay back. A dishonest-type E should never renege in the first few plays, then should begin using a "mixed strategy"—a probabilistic combination of "pure" strategies—with an increasing probability of reneging each period. (In the eighth period, a dishonest E should certainly renege.) Once the dishonest E begins mixing, B should begin mixing too, lending with a probability of .64. The dishonest E plays a mixed strategy because E wants to maintain a reputation for possibly being honest by paying back loans for as many periods as possible. (Once she reneges, her reputation is shot, and she gets no more loans.) She does not want to renege in the same period every time, or the B's will catch on and withhold loans in that period.

In the first 20 or so 8-period games, subjects deviated from the game-theoretic equilibrium in decision-theoretic ways. For instance, when E reneges in an early period, Bs should learn that she is dishonest and refuse to make loans. Instead, many Bs would lend in later periods, and E would renege again. Thinking decision theoretically, they regarded these bad loans

4. Actually, there are many sequential equilibria. The one we pick out is the only one that passes the "intuitive criterion," discussed below.

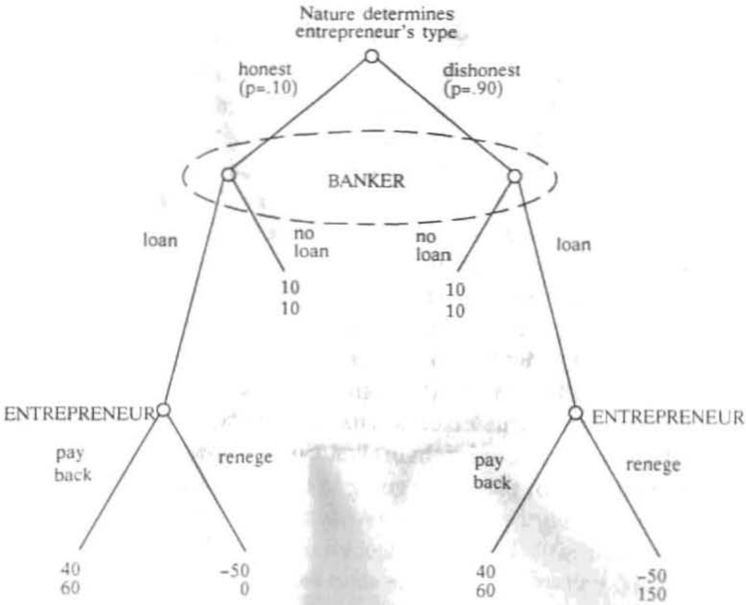

Figure 13.2. The banker-entrepreneur reputation game (Camerer and Weigelt 1988). Note that the upper (lower) payoff is the banker's (entrepreneur's).

as gambles that might be paid back because E's choice is unpredictable. Thinking game theoretically—knowing E's payoffs and assuming her rationality—makes the choice predictable: later loans will not be paid back. (They were not.)

In newer experiments, we measured subjects' estimates of the probability of having a loan paid back in each period of each game. Probabilities were biased in familiar ways (e.g., extreme probabilities are used too often; see Lichtenstein, Fischhoff, and Phillips 1982). The most striking finding is that other E players were better than B players at estimating whether a particular E would pay a loan back, though the Bs and Es had exactly the same objective data available to them when they made estimates. It appears that being in the same game-theoretic role as another player improves insight, helping subjects learn mutual rationality, just as playing future periods helps subjects backward induct.

After many 8-period games, play did converge remarkably closely to the equilibrium predictions. There was one persistent systematic deviation: dishonest E subjects did not renege as early in the game as predicted. This deviation makes sense if players believe that a fraction of dishonest Es (around 16 percent) will always behave honestly. This is further evidence of cooperation, as discussed above, because any E can do better by not behaving honestly but extra honest players increase profits for everyone.

Refinements of Nash Equilibrium

In games with many equilibria, the need to judge others' choices correctly is especially acute. Theorists have developed rules for evaluating the logic of different equilibria, called "refinements."

Many refinements have been described for incomplete-information games like the banker-entrepreneur game. Another example is the education game between students and employers, shown in figure 13.3 (adapted from Cho and Kreps 1987). It is a signaling game in which an informed party chooses a signal that another party sees and responds to.

In the education game, students are of two types, bright and dim. (Assume $p[\text{bright}] = .6$ and $p[\text{dim}] = .4$.) A student observes her type, then decides whether to attend college or not (the signal). Employers see whether a student went to college or not and decide whether to hire her, but employers do not know whether the student is bright or dim. Students get a payoff of 2 units of utility for getting a job, plus 1 for going to college if they are bright or 1 for

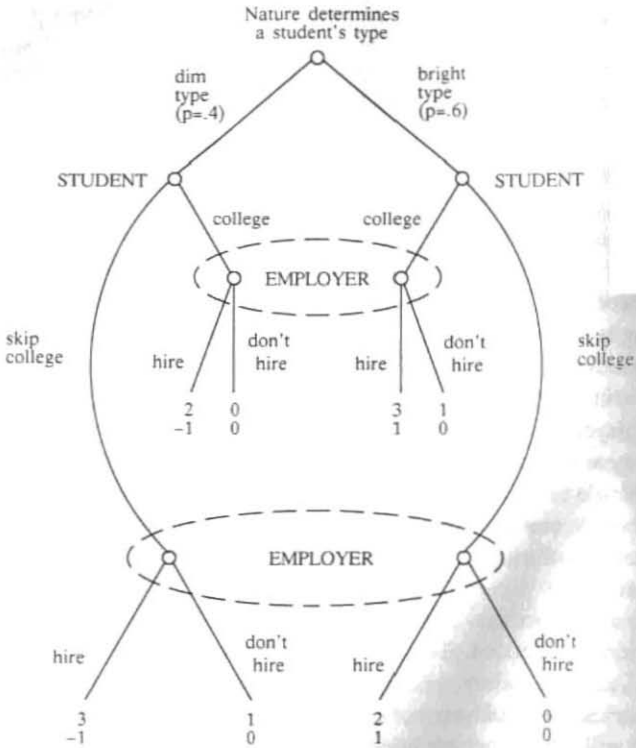

Figure 13.3. The education game (adapted from Cho and Kreps 1987). Note that the upper (lower) payoff is the student's (employer's).

avoiding college if they are dim. Employers get 1 for hiring a bright student and lose 1 for hiring a dim one.

There are two pure-strategy sequential equilibria in the education game. In one equilibrium, both bright and dim students skip college. (This is called a "pooling" equilibrium because both types choose the same strategy and enter a common pool.) Since both types skip college, employers using Bayes's rule realize that $p(\text{bright} \mid \text{no college}) = p(\text{bright}) = .6$. They then hire, earning an expected utility of $.6(1) + .4(-1) = .2$ (instead of not hiring and earning 0).

For this to be a Nash equilibrium, employers must not hire a student who went to college. (Otherwise, skipping college would not be a best response for a bright student.) For the equilibrium to be sequential (a refinement of Nash), employers must have beliefs that make not hiring a college skipper rational. To make not hiring rational, employers must believe that $p\,(\text{bright}|\text{college}) < .5$ (then the expected utility of hiring is less than zero, the certain value of hiring).

Since $p\,(\text{bright}) = .6$, the employer is acting as if going to college lowers the perceived probability that the student is bright. This inference seems backward. Dim students cannot possibly do any better by leaving the "equilibrium path" (where they earn 3) and going to college. Bright students might do better (they earn 2 in the equilibrium and might earn 3 by switching). A logical refinement called the "intuitive criterion" (Cho and Kreps 1987) states that $p\,(\text{bright}|\text{college})$ should be one if bright students might benefit from college and dim students certainly will not. But then it is optimal for employers to hire if they observe college (earning them 1 for sure), which makes going to college a best response for bright types, shattering the equilibrium. The resulting equilibrium does pass the intuitive criterion: all students go to college and get hired; students who skip college are thought probably to be dim and do not get hired.

In experimental tests of such games, the judgments of others' choices are often decision theoretic rather than game theoretic. For instance, undergraduates in a game-theory course made the choices shown in table 13.1. Subjects were asked what they would do at every possible point in the game. If dim, only six of 43 subjects would go to college; if bright, all go to college. As employers, 41 of 43 would hire college goers, but only seven of 43 would hire college skippers. Their logic is not game theoretic because, when dim students decide to skip college, they are apparently not thinking through what rational employers would do—not hire—if faced with a college skipper. (In this experiment, their choices as dim students are inconsistent with their own choices as employers.)

In more thorough experiments (Brandts and Holt 1987), dim types skip college 70 percent of the time. A natural explanation is that subjects are making "maximin" choices, maximizing their worst possible outcome. (In the education game, dim types can get at least 1 by skipping college.) Maximin

TABLE 13.1
EDUCATION GAME RESULTS

	Choice	
	College	Skip
Type:		
Dim	6	37
Bright	43	0
Employer choice:		
Hire	41	7
Not hire	2	36

Note. $N = 43$.

choices are defensible in many contexts (when the sum of all payoffs is constant, a Nash equilibrium necessarily consists of maximin choices), but not here; players choosing according to maximin are regarding choices by others as less predictable (or more malicious) than they truly are. Banks, Camerer, and Porter (1988) reported related results but found that choices were not always consistent with the maximin rule.

SOME PARALLELS WITH BEHAVIORAL DECISION THEORY

My general approach builds behavioral game theory up around the special features of games, deliberately distinguishing it from behavioral decision theory. The two have much in common as well.

Context

Seemingly innocuous changes in context affect the outcomes of games. For instance, experiments and surveys described above suggest that perceived fairness of decisions is sensitive to the method by which decision-making power was granted.

Schelling (1960) noted that many games with several equilibria have "focal points"—"psychologically prominent" equilibria—that are suggested by context and circumscribed by culture. An example appeared in *Games* magazine in November 1988. Readers were shown cartoons of nine celebrities and instructed to "vote" for one celebrity. If you voted for the celebrity receiving the most votes, you became eligible for prizes. The reader's job was to guess what people would do, knowing that those people would be guessing what others would do, ad infinitum.

The *Games* game has nine equilibria in pure strategies, one for each celebrity: those who think that the largest number of people will vote for Pee Wee Herman will vote for him too; those who think that Shirley MacLaine will receive the largest number of votes will vote for her; and so on. Voters man-

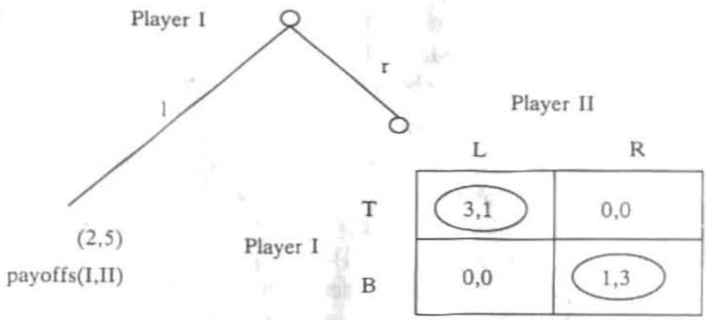

Figure 13.4. A game in which context matters (van Damme 1987).

aged to achieve a remarkable consensus: Bill Cosby won the actual contest (1,489 votes), edging out Lee Iacocca (1,155 voters) and distancing the seven lesser-known celebrities (2,639 votes combined). The same game in another context—with strategies labeled 1–9 or A–I—would not have produced such consensus.

The celebrity game is a "coordination game"; people are better off if they can coordinate their choices, doing what others do. In coordination games, context sets historical precedent, which has great influence (see, e.g., van Huyck, Battalio, and Beil 1988). For instance, driving on the right side of the road is a coordination equilibrium in America (now enforced by law) because of a convention established centuries ago. In England, a different history created a different convention, driving on the left.[5]

A final example of context is shown in figure 13.4 (see van Damme 1987). Player I moves first, choosing l or r. If r is chosen, players I and II play a simultaneous move game (shown in matrix form). The matrix game has two Nash equilibria in pure strategies (circled): (T, L), yielding (3, 1), and (B, R), yielding (1, 3).

If the matrix game were played separately, there is no reason to think that either equilibrium is more likely (unless labeling of strategies induces a focal point). But playing the matrix game after player I moved r is different. The matrix game is now circumscribed by context: by giving up her payoff of 2 (from [2, 5]) and taking her chances on the matrix game, player I is hinting to player II that they should play the equilibrium that gives her a payoff of 3 (i.e., [T, L]). Why else would she give up a certain payoff of 2? (This reasoning is

5. The American convention was established by farmers driving large teams of horses to market. They sat on the left rear horse so that they could lash the team with a whip, right handed. Since they were sitting on the left, accidents were best avoided if other teams passed on the left; they drove on the right. English drivers sat up on smaller carriages with a load behind them. A whip lashed right handed would get caught in the load if drivers sat on the left, so they sat on the right. Drivers passed on the right. Historical context matters: on an otherwise identical planet with more left-handed drivers, Americans would drive on the left, the English on the right.

called "forward induction": moves are assumed to tell players something about intentions in future subgames.) In this game, foregone choices provide a context; the context might matter because foregoing a choice means something.

Feedback and Learning

The process of learning is crucial and almost completely neglected in game theory. An exceptional model is Harsanyi's "tracing procedure" (1975; cf. Fudenberg and Kreps 1988). Players begin by guessing the probabilities with which other players choose strategies. Then they choose strategies that maximize their own expected utility given their guesses. This procedure yields an optimal strategy for each player; guesses are revised by shifting probability onto each player's optimal strategy, and the procedure is repeated until an equilibrium emerges. The tracing procedure was proposed as a description of the way players think before they play the game, but it seems even more useful as a model of learning across plays.

Behavioral decision theory suggests a simple model of learning: people respond to feedback by changing strategies until they reach a point from which they can do no better—the hallowed grounds of equilibrium. Convergence thus requires three ingredients: feedback must be clear, immediate, and repeated; subjects must interpret feedback correctly, realizing that they can do better; and subjects must change in the correct direction.[6] The most dramatic learning occurs in experiments that satisfy all three conditions. Recall the strategic voting experiment of Eckel and Holt (1989; fig. 13.1 above). In the second stage, A voters get clear feedback about the implications of their first-stage vote. Since their worst outcome is elected, subjects realize that a change can only help. Since there is only one way for them to change (by voting for {B, C} instead of {A, B}), they cannot help but change in the right direction.

In most of the sequential bargaining (pie-splitting) experiments, learning conditions were poor, and convergence was too. Subjects got no feedback about how much better they would do if they made more aggressive offers because initial offers were rarely rejected. When feedback was provided by playing subgames separately (in Harrison and McCabe's experiment), subjects converged remarkably close to perfect equilibrium.

Lack of feedback has an interesting effect in a game studied by Schotter, Weigelt, and Wilson (1988). Their game is depicted in extensive form in figure 13.5 and in normal form in table 13.2.

Player 1 moves first. There are two Nash equilibria, (L, l) and (R, r). (Only the [R, r] equilibrium survives logical refinements.) If player 1 regards player 2's choice as a random variable rather than an optimal action deducible by introspection, she may choose L because it guarantees a payoff of 4 (it is the maximin strategy). But r dominates l for player 2. Player 1 should realize

6. Hilly Einhorn's influence here (e.g., Einhorn 1980) should be obvious.

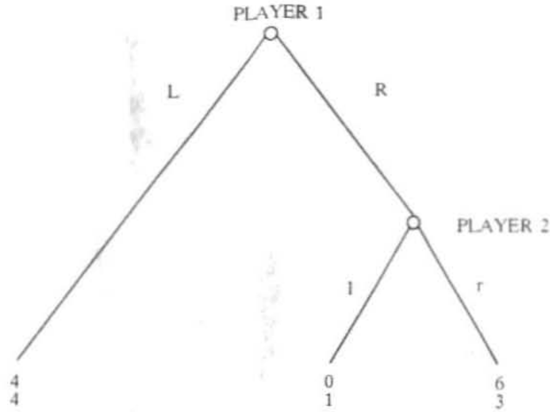

Figure 13.5. A transparency game (Schotter, Weigelt, and Wilson 1988). Note that the upper (lower) payoff is player 1's (player 2's).

TABLE 13.2
A TRANSPARENCY GAME

	Player 2	
	l	r
Player 1:		
L	4, 4	4, 4
R	0, 1	6, 3

Source. Schotter, Weigelt, and Wilson (1988).

this and choose R, earning 6 instead of 4. When the game was shown in tree form (fig. 13.5), player 1 subjects chose R 98 percent of the time. Shown the game in matrix form (table 13.2), player 1 subjects chose R only 44 percent of the time. Presenting the game in tree form seems to make it transparent that r is a dominant strategy for player 2 (cf. Keller 1985).

In this experiment, subjects were told only their own payoffs at the end of each period. When player 1 subjects chose L, they learned that they got 4, but they did not know whether player 2 chose l or r. I suspect that feedback is crucial for generating convergence in this matrix game: without knowing more than just their payoff, player 1 subjects could not be sure that R was a better choice than L. (Feedback on player 2's choices did cause player 1 to switch from L to R in classroom experiments.)

Levels of Explanation

An important finding in behavioral decision theory is that simple linear models may explain complicated judgments "paramorphically," without de-

TABLE 13.3
A GAME WITH A MIXED-STRATEGY EQUILIBRIUM

	Column Player				Frequency	
	1	2	3	J	Predicted	Actual
Row player:						
1	−5, 5	5, −5	5, −5	−5, 5	.20	.221
2	5, −5	−5, 5	5, −5	−5, 5	.20	.215
3	5, −5	5, −5	−5, 5	−5, 5	.20	.203
J	−5, 5	−5, 5	−5, 5	5, −5	.40	.362
Frequency:						
Predicted	.20	.20	.20	.40		
Actual	.226	.179	.169	.426		

Source. O'Neill (1987).

scribing the detailed process of judgment very well (Einhorn, Kleinmuntz, and Kleinmuntz 1979). Models that work well at one level (overall judgment) work poorly at a less-aggregated level (judgment process). A similar phenomenon occurs with predictions of mixed-strategy play. A mixed strategy is a random choice (or "probabilistic mixture") of pure strategies.

Consider the game in table 13.3 (used by O'Neill 1987). There is no pair of strategies that are mutual best responses. However, if the column player chooses the strategies (1, 2, 3, J) with probabilities (.2, .2, .2, .4), then the row player's expected utility is the same for all strategies (−1). A mixed strategy with the same probabilities is therefore a weak best response to the column player (no other strategy is better; no other strategy is worse either).

O'Neill (1987) found that subjects played each strategy with an overall frequency remarkably close to that predicted (the data are shown in table 13.3), but the predictions are much less impressive at the individual level. (Brown and Rosenthal [1987] reject the predictions for a third of the subjects at $p < .05$.) Like paramorphic models, the mixed-strategy prediction works well at one level but fails at another.

In banker-entrepreneur reputation games (Camerer and Weigelt 1988), relative frequencies of choices were remarkably similar to the predicted probabilities overall, but individuals chose different patterns, which were often deterministic rather than mixed (e.g., pay back until period 5, then renege). Since different people chose different deterministic patterns, a player facing a randomly chosen opponent confronted an unpredictable strategy choice, which was effectively mixed.

Bull, Schotter, and Weigelt (1987) observed the opposite phenomenon in their experimental "tournaments." A tournament is a labor contract in which players choose effort levels (higher effort is more costly) and output is the sum of effort and a random variable. The player with the largest output wins a fixed prize (e.g., tenure or the company presidency). In their experiments, there is a unique pure-strategy equilibrium—subjects should choose the same effort level each time. Average effort was remarkably close to the level predicted, but individual effort varied dramatically over time and across people (perhaps because subjects tried to outguess the random variable, as in probability matching). The Nash equilibrium prediction was quite accurate in the aggregate but cannot explain individual variation. People cannot generate random sequences easily (e.g., Baddeley 1966), though they can learn to do so with extensive feedback (Neuringer 1986). Thus, it is not surprising that people do not mix strategies randomly (for almost half O'Neill's subjects, choices depended on their own previous choices). Nonrandom mixing is troublesome for game theory because it changes equilibrium strategies if subjects can detect it. It would be useful to know if they can detect it.

CONCLUSION

My argument is that game theory relies on descriptively inadequate assumptions of the two features that distinguish games from decisions. The first feature is that games yield a payoff to each player. If players care about others' payoffs, predictions based purely on self-interest will be wrong. A wide variety of data suggest that people typically do prefer fair payoffs (often equal ones) but that their concern for fairness is context dependent. People also try to cooperate by making personal sacrifices to maximize joint gains.

The second feature is that, in games, players must judge the choices that others will make (including themselves in the future). The special assumption of equilibrium analysis in game theory is that such judgments are made by considering how others will behave if they are rational. Data suggest that people do not consider others. They are also myopic in anticipating their own future choices.

Of course, we should not abandon game theory simply because people violate it. Equilibrium predictions provide a handy, precise target that people move toward. Subjects usually begin experiments playing fairly, myopically, and decision theoretically. Gradually, they learn to accept unfair outcomes, plan ahead, and expect rational choices by others (if the equilibrium dictates that they should). A central question for *behavioral* game theory is learning conditions for convergence.

The natural hypothesis is that clear, informative feedback and the ability to adapt are necessary for convergence. Most of the experiments surveyed are especially conducive to learning, and convergence usually occurs (but never immediately). More experiments that are less conducive to learning (like the sequential bargaining experiments in which unreached subgames are not played) would be useful to see if convergence fails.

Then we must ask when people converge to equilibria in the natural world. Is the world more like the first period of an experiment or the last? The answer is surely mixed. Novice negotiators, first-time home buyers, and newlyweds are probably initially myopic in sequential bargaining, like inexperienced subjects. Veterans of painful strikes or divorces have lived through subgames (as experienced subjects have); they are probably not so myopic in new games. Practice may help too: novice salesmen rehearse the closing of a sale (imagining the end of their sequential bargaining with customers before it begins); hopeful lovers might do the same before a date.

If subjects can learn, an important question is how well learning in one setting transfers to another setting. Is behavior in the first period of an experiment like that in the last period of a similar experiment? There are good reasons to be pessimistic about transfer (on the "winner's curse," see, e.g., Kagel and Levin 1986, 909–10), but there are too few data to say anything more.

Broader data would be especially useful in developing behavioral game theory. In virtually all the experiments described in this chapter, researchers collected only the data needed to test the normative theory that they considered (choices, typically). It is easy to collect lots of other data, like judgments of what others will do (e.g., Selten and Stoecker 1986), process measures, or protocols. From those data, we can make game theory more behavioral, and better.

REFERENCES

Andreoni, J. 1988. Why free ride? Strategies and learning in public goods experiments. Working paper. University of Wisconsin, Department of Economics.

Aumann, R. J. 1985. What is game theory trying to accomplish? In *Frontiers of economics*, ed. Kenneth J. Arrow and Seppo Honkapohja. Oxford: Blackwell.

———. 1987. Correlated equilibrium as an expression of Bayesian rationality. *Econometrica* 55:1–18.

Baddeley, A. D. 1966. The capacity for generating information by randomization. *Quarterly Journal of Experimental Psychology* 18:119–28.

Banks, J., C. F. Camerer, and D. Porter. 1988. Experimental tests of Nash refinements in signaling games. Working paper. University of Pennsylvania, Department of Decision Sciences.

Bar-Hillel, M., and M. Yaari. 1987. Judgments of justice. Working paper. Hebrew University, Department of Economics.

Bazerman, M. H., and J. S. Carroll. 1987. Negotiator cognition. In *Research in organizational behavior*, ed. B. M. Staw and L. L. Cummings. Greenwich, Conn.: JAI.

Binmore, K. 1987. Modeling rational players. *Economics and Philosophy* 3:179–214.

Binmore, K., A. Shaked, and J. Sutton. 1985. Testing noncooperative bargaining theory: A preliminary study. *American Economic Review* 75:1178–80.

———. 1988. A further test of noncooperative bargaining theory: Reply. *American Economic Review* 78:837–39.

Brandts, J., and C. A. Holt. 1987. An experimental test of equilibrium dominance in signaling games. Working paper. University of Virginia, Department of Economics.

Brown, J. N., and R. W. Rosenthal. 1987. Testing the minimax hypothesis: A reexamination of O'Neill's game experiment. Working paper. State University of New York, Stonybrook, Department of Economics.

Brubaker, E. R. 1982. Sixty-eight percent free revelation and thirty-two percent free ride: Demand disclosures under varying conditions of exclusion. In *Research in experimental economics*, vol. 2, ed. V. L. Smith. Greenwich, Conn.: JAI.

Bull, C., A. Schotter, and K. Weigelt. 1987. Tournaments and piece rates: An experimental study. *Journal of Political Economy* 95:1–33.

Camerer, C., G. Loewenstein, and M. Weber. 1989. The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy* 97:1232–54.

Camerer, C., and K. Weigelt. 1988. Experimental tests of a sequential equilibrium reputation model. *Econometrica* 56:1–36.

Cho, I.-K., and D. M. Kreps. 1987. Signaling games and stable equilibria. *Quarterly Journal of Economics* 102:179–221.

Coase, R. 1960. The problem of social cost. *Journal of Law and Economics* 3:1–44.

Colman, A. 1983. *Game theory and experimental work*. London: Pergamon.

Coursey, D. L., E. Hoffman, and M. L. Spitzer. 1987. Fear and loathing in the Coase theorem: Experiments involving physical discomfort. *Journal of Legal Studies* 16:217–48.

Coursey, D. L., R. M. Isaac, and V. L. Smith, 1984. Natural monopoly and contested markets: Some experimental results. *Journal of Law and Economics* 27:91–113.

Dawes, R. M. 1980. Social dilemmas. *Annual Review of Psychology* 31:169–93.

Dawes, R. M., J. McTavish, and H. Shaklee. 1977. Behavior, communication, and assumptions about other people's behavior in a common dilemma situation. *Journal of Personality and Social Psychology* 35:1–11.

Dawes, R. M., and J. Orbell. Social dilemmas. In *Progress in applied social psychology*, vol. 1, ed. G. Stephenson and J. Davis. Chichester: Wiley.

Dawes, R. M., and R. H. Thaler. 1988. Anomalies: Cooperation. *Journal of Economic Perspectives* 2:187–97.

Dawes, R. M., A. J. C. van de Kragt, and J. M. Orbell. 1988. Not me or thee but we: The importance of group identity in eliciting cooperation in dilemma situations: Experimental manipulations. *Acta Psychologica* 68:83–97.

Eckel, C. and C. A. Holt. 1989. Strategic voting in agenda-controlled committees. *American Economic Review* 79:763–73.

Einhorn, H. J. 1980. Learning from experience and suboptimal rules in decision making. In *Cognitive processes in choice and decision behavior*, ed. T. S. Wallsten. Hillsdale, N.J.: Erlbaum.

Einhorn, H. J., D. N. Kleinmuntz, and B. Kleinmuntz. 1979. Linear regression and process tracing models of judgment. *Psychological Review* 86:465–85.

Forsythe, R., J. Kennan, and B. Sopher. 1987. An experimental analysis of bargaining and strikes with one sided private information. Working paper no. 87–4. University of Iowa, Department of Economics.

Forsythe, R., T. Palfrey, and C. Plott. 1982. Asset valuation in an experimental market. *Econometrica* 50:537–67.

Fudenberg, D., and D. Kreps. 1988. A theory of learning, experimentation, and equilibrium in games. Working paper. Stanford University, Graduate School of Business, July.

Güth, W., R. Schmittberger, and B. Schwarze. 1982. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization* 3:367–88.

Güth, W., and R. Tietz. 1986. Auctioning ultimatum bargaining positions—how to act if rational decisions are unacceptable? In *Current issues in West German decision research*, ed. R. W. Scholz. Frankfurt: Lang.

———. 1988. Ultimatum bargaining for a shrinking cake—an experimental analysis. In *Bounded rational behavior in experimental games and markets*, Lecture Notes in Economics and Mathematical Systems, vol. 314, ed. R. Tietz, W. Albers, and R. Selten. Berlin: Springer.

Harrison, G., and J. Hirshleifer. In press. An experimental evaluation of weakest-link/best-shot models of public goods. *Journal of Political Economy.*

Harrison, G., and K. McCabe. 1988. Testing bargaining theory in experiments. Working paper. University of Western Ontario, Department of Economics.

Harrison, Glenn, and M. McKee. 1985. Experimental evaluation of the Coase theorem. *Journal of Law and Economics* 28:653–70.

Harsanyi, J. C. 1967–68. Games with incomplete information played by "Bayesian" players. I–III. *Management Science* 14:159–82, 320–34, 486–502.

———. 1975. The tracing procedure: A Bayesian approach to defining a solution for *n*-person noncooperative games. *International Journal of Game Theory* 4:61–94.

———. 1982. Subjective probability and the theory of games: Comments on Kadane and Larkey's paper, and rejoinder to reply. *Management Science* 28:120–24, 125.

Hoffman, E., and M. Spitzer. 1982. The Coase theorem: Some experimental tests. *Journal of Law and Economics* 25:73–98.

———. 1985. Entitlements, rights, and fairness: An experimental examination of subjects' concepts of distributive justice. *Journal of Legal Studies* 14:259–97.

———. 1986. Experimental tests of the Coase theorem with large bargaining groups. *Journal of Legal Studies* 15:149–71.

Isaac, R. M., K. F. McCue, and C. R. Plott. 1985. Public goods provision in an experimental environment. *Journal of Public Economics* 26:51–74.

Kadane, J. B., and P. D. Larkey. 1982. Subjective probability and the theory of games, and reply. *Management Science* 28:113–20, 124.

Kagel, J. H., and D. Levin. 1986. The winner's curse and public information in common value auctions. *American Economic Review* 76:894–920.

Kahan, J. P., and A. Rapoport. 1984. *Theories of coalition formation.* Hillsdale, N.J.: Erlbaum.

Kahneman, D., J. L. Knetsch, and R. H. Thaler, 1986a. Fairness and the assumptions of economics. *Journal of Business* 59:S285–S300.

———. 1986b. Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review* 76:728–41.

Keller, L. R. 1985. The effects of problem representation on the sure-thing and substitution principles. *Management Science* 31:738–51.

Kim, O., and M. Walker. 1984. The free rider problem: Experimental evidence. *Public Choice* 43:3–24.

Kramer, R. M., and M. B. Brewer. 1986. Social group identity and the emergence of cooperation in resource conservation dilemmas. In *Experimental social dilemmas,* ed. H. Wilke, D. Messick, and C. Rutte, 205–34. Frankfurt am Main: Lang.

Kunreuther, H. 1986. Comments on Plott and on Kahneman, Knetsch, and Thaler. *Journal of Business* 59:S329–S335.

Lichtenstein, S., B. Fischhoff, and L. D. Phillips. 1982. Calibration of probabilities: The state of the art to 1980. In *Judgment under uncertainty: Heuristics and biases,* ed. D. Kahneman, P. Slovic, and A. Tversky. Cambridge: Cambridge University Press.

Loewenstein, G., L. Thompson, and M. Bazerman. 1989. Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology* 57:426–41.

Marwell, G., and R. E. Ames. 1979. Experiments on the provision of public goods. I. Resources, interest, group size, and the free-rider problem. *American Journal of Sociology* 84:1335–60.

———. 1980. Experiments on the provision of public goods. II. Provision points, stakes, experience, and the free-rider problem. *American Journal of Sociology* 85:926–37.

———. 1981. Economists free ride, does anyone else? Experiments on the provision of public goods. IV. *Journal of Public Economics* 15:295–10.
McKelvey, R. D., and P. C. Ordeshook. 1987. A decade of experimental research on spatial models of elections and committees. Social Science working paper no. 657. California Institute of Technology.
Messick, D. M., and M. B. Brewer. 1983. Solving social dilemmas: A review. In *Review of personality and social psychology*, ed. L. Wheeler and P. Shaver, vol. 4, 11–44. Beverly Hills: Sage.
Mumpower, J. L. 1988. Heuristics, biases, and optimality in negotiations. Working paper. State University of New York, Administration and Policy.
Myerson, R. B. 1986. Negotiation in games: A theoretical overview. In *Uncertainty, information, and communication: Essays in honor of Kenneth J. Arrow*, vol. 3, ed. W. Heller, R. Starr, and D. Starrett, 3–24. New York: Cambridge University Press.
Nash, J. 1951. Non-cooperative games. *Annals of Mathematics* 54:286–95.
Neelin, J., H. Sonnenschein, and M. Spiegel. 1988. A further test of noncooperative bargaining theory: Comment. *American Economic Review* 78:824–36.
Neuringer, A. 1986. Can people behave "randomly"? The role of feedback. *Journal of Experimental Psychology: General* 115:62–75.
Ochs, J., and A. E. Roth. 1989. An experimental study of sequential bargaining. *American Economic Review* 79:355–84.
O'Neill, B. 1987. Nonmetric test of the minimax theory of two-person zero-sum games. *Proceedings of the National Academy of Sciences, U.S.A.* 84:2106–9.
Orbell, J. M., A. J. C. van de Kragt, and R. M. Dawes. In press. Explaining discussion-induced cooperation in social dilemmas. *Journal of Personality and Social Psychology*.
Prudencio, Y. C. 1982. The voluntary approach to externality problems: An experimental test. *Journal of Environmental Economics and Management* 9:213–28.
Radner, R., and A. Schotter. 1987. The sealed-bid mechanism: An experimental study. Working paper no. 87-41. New York University, C. V. Starr Center for Applied Economics.
Rapoport, A. 1988. Provision of step-level public goods: Effects of inequality in resources. *Journal of Personality and Social Psychology* 54:432–40.
Roth, A. E. 1987. Bargaining phenomena and bargaining theory. In *Laboratory experimentation in economics: Six points of view*, ed. A. E. Roth. Cambridge: Cambridge University Press.
Rubinstein, A. 1982. Perfect equilibrium in a bargaining model. *Econometrica* 50: 97–109.
———. 1988. Comments on the interpretation of game theory. Discussion paper no. TE/88/181. London School of Economics.
Schelling, T. 1960. *The strategy of conflict*. Cambridge, Mass.: Harvard University Press.
Schotter, A., K. Weigelt, and C. Wilson. 1988. Strategic choice and presentation effects: Some experimental evidence. Working paper. New York University. C. V. Starr Center for Applied Economics.
Selten, R. 1987. Equity and coalition bargaining in experimental games. In *Laboratory experimentation in economics: Six points of view*, ed. A. E. Roth. Cambridge: Cambridge University Press.
Selten, R., and R. Stoecker. 1986. End behavior in sequences of finite prisoner's dilemma supergames. *Journal of Economic Behavior and Organization* 7:47–70.
Simmons, R. T., R. M. Dawes, and J. M. Orbell. 1983. An experimental comparison of the two motives for not contributing to a public good: Desire to free ride and fear

of being gypped. Working paper. University of Oregon, Department of Psychology.
Smith, V. L. 1981. An empirical study of decentralized institutions of monopoly restraint. In *Essays in contemporary fields of economics*, ed. G. Horwich and J. Quirk. West Lafayette, Ind.: Purdue University Press.
Stroebe, W., and B. S. Frey. 1982. Self-interest and collective action: The economics and psychology of public goods. *British Journal of Social Psychology* 21:121–37.
Sutton, J. 1986. Non-cooperative bargaining theory: An introduction. *Review of Economic Studies* 53:709–24.
Thaler, R. 1988. Anomalies: The ultimatum game. *Journal of Economic Perspectives* 2:195–206.
van Damme, Eric. 1987. Stable equilibria and forward induction. Working paper. University of Bonn.
van de Kragt, A. J. C., J. M. Orbell, and R. M. Dawes. 1983. The minimal contributing set as a solution to public goods problems. *American Political Science Review* 77:112–22.
van de Kragt, A. J. C., J. M. Orbell, and R. M. Dawes with S. R. Braver and L. A. Wilson II. 1986. Doing well and doing good as ways of resolving social dilemmas. In *Experimental social dilemmas*, ed. H. Wilke, D. Messick, and C. Rutte. Frankfurt am Main: Lang.
van Huyck, J. B., R. C. Battalio, and R. O. Beil. 1988. Strategic uncertainty, equilibrium selection principles, and coordination failure in average opinion games. Working paper no. 88-20. Texas A&M, Department of Economics.
Wilson, R. 1985. Reputations in games and markets. In *Game-theoretic models of bargaining*, ed. A. E. Roth. Cambridge: Cambridge University Press.
Yaari, M., and M. Bar-Hillel. 1984. On dividing justly. *Social Choice and Welfare* 1:1–24.